Commentary

Why Have We Failed to Find the Low Penetrance Genetic Constituents of Common Cancers?

Neil E. Caporaso[1]

Genetic Epidemiology Branch, National Cancer Institute, Rockville, Maryland 20892

Introduction

Despite widespread enthusiasm for the candidate gene approach to identify common but "low penetrance" susceptibility genes that predict cancer risk (1), in the aggregate, association studies involving complex diseases have yielded a scant harvest (2). One review concluded recently that only 6 of 166 disease associations that were studied three or more times were replicated consistently (3). Another systematic review of genetic polymorphisms and breast cancer could not exclude a potential role for a variety of candidates but neither could it verify significant risk associated with any specific gene (4). Earlier studies, as well as meta-analyses involving putative susceptibility genes and lung cancer have been inconclusive. Therefore, uncertainty as to whether polymorphic variation in the *MPO*[2] gene alters susceptibility to lung cancer is more typical rather than unique, and we need to examine the issue in a broad context. This provides an opportunity to consider why well-conducted studies such as the reports by Feyler *et al.* (5) and Xu *et al.* (6) involving the relationship of the MPO gene polymorphism and lung cancer arrive at seemingly opposing conclusions.

I will frame the discussion of the genetics of complex disease as pertains to cancer in general and MPO and lung cancer in particular around the following questions.

Is there sound evidence that a hereditary component of common malignancies exist?

If inherited susceptibility to common cancers exists, is it important to identify?

What does the current evidence indicate?

If finding genetic susceptibility factors is important, why after more than a decade of work has it not clearly identified? What can be done to resolve the question?

Evidence for an Influence of Heredity in Common Cancers

The preponderance of evidence from diverse approaches supports at least some hereditary component for virtually all of the major malignancies. Approaches include enumerating cancer in relatives of cases compared with the relatives of controls and twin studies that show enhanced concordance in monozygotic compared with dizygotic twins (7, 8). The high penetrance genes that create striking familial aggregations of cancer (*i.e.,*

the *BRCA* mutations and breast cancer) only explain a fraction of the genetic component of risk.

The earliest studies to evaluate candidate genes and tobacco-related cancers began in the pregenotyping era (*i.e.,* before 1998). Probe drugs corresponding to potentially carcinogenic compounds metabolized by the genes of interest were administered and "metabolic phenotypes" constructed based on the distribution of recovered metabolites; genotypes were inferred. Of the first four metabolic polymorphisms studied by this approach (9–12), only one association can be considered well established. Mechanistic (13) and population data (14) are reasonably concordant, and establish that NAT2 "slow acetylators" exhibit increased risk for bladder cancer. Evidence that polymorphisms in CYP1A1, GSTM1, or CYP2D6 influence susceptibility to lung cancer remains controversial. For the first two, mechanistic evidence is convincing (15), but meta-analyses now including a substantial number of studies exhibit only weak evidence for a small effect (16–18). For CYP2D6, both mechanistic evidence and population data are null to weakly suggestive (19).

Why Is It Important to Identify Low Penetrance Genes That Contribute to Cancer?

Given that the environment plays a dominant role in most common cancers (20, 21), why is identifying a "minor" genetic component important? The history of epidemiology indicates that progress in prevention (*i.e.,* public health) and treatment (*i.e.,* clinical) of cancer can occur without knowledge of the specific mechanism (or genes) that contribute to susceptibility. However, for many malignancies, substantive advances in both etiologic understanding (*i.e.,* brain and prostate) and in therapy (*i.e.,* lung) have been elusive. Given the slow progress, improved insight from understanding how genes promote or block cancer can be crucial. Moreover, even a modest contribution of a common genetic factor would imply a substantial attributable risk (22). A clear understanding of one or more susceptibility genes could provide improved mechanistic insight, clues to exposure mechanisms (*i.e.,* a gene clearly related to susceptibility may suggest a specific etiologic agent), or may suggest a target for chemoprevention or other interventions. Even more importantly, such a finding can be seen as key to elucidating a fundamental question about cancer. Given that on the population level, most cancer depends on environmental factors, but on a cellular level, genetic lesions characterize cancer, understanding how these two processes intersect will be of fundamental importance. This is true even if in most instances, hereditary susceptibility genes are not the same as the genes somatically mutated.

What Does the Evidence Suggest and Do the Classic Criteria for Evaluation of Causality Help?

What can we conclude from the published *MPO*/lung cancer studies, and what are the implications for the broader literature?

One starting point is to consider the major criteria for the evaluation of causality (23, 24) that include biological plausibility, strength and specificity of the association, temporality, and dose response.

Biological Plausibility. MPO is an enzyme that generates reactive oxygen species and thereby contributes to bactericidal activity in white cells. The cost of enhanced killing by the more active forms of MPO is increased inflammation, and the balance of polymorphic gene variants that alter activity likely reflect selection pressure in different populations. In a manner similar to other polymorphic Phase I enzymes (*i.e.,* CYP1A1), MPO converts polycyclic hydrocarbons such as benzo(a)pyrene to DNA-binding intermediates (25) potentially contributing to pulmonary carcinogenesis, as well as to other conditions such as arteriosclerosis and Alzheimer's (26). A G-463-A variant in the promoter region of MPO alters an SP1 binding site and results in decreased transcriptional activity. One report suggests the action of MPO is additionally catalyzed by nicotine, suggesting the possibility of enhanced effect in smokers (27). The evidence for biological plausibility is generally supportive.

Strength of Association. Before the studies in the current issue, the hypothesis of an association between an MPO variant and lung cancer has been addressed in six published studies: London *et al.* (28), LeMarchand *et al.* (29), Cascorbi *et al.* (30), Schabath *et al.* (31), Misra *et al.* (32), and Kantarcci *et al.* (33). A detailed quantitative treatment of the data from all of the previous studies involving MPO and lung cancer exceeds the scope of this article but considering the six earlier published studies, a summary risk ratio is 0.92 (95% confidence interval, 0.75–1.04) for heterozygotes and 0.66 (0.36–0.78) for AA homozygotes with borderline not significant heterogeneity. Three of the six studies are nominally "significant," and five of the six studies exhibit an overall effect in the direction hypothesized. When the small positive study by Feyler *et al.* (5) and a large convincingly null study by Xu *et al.* (6) are included, the summary risk ratios lose statistical significance, whereas heterogeneity becomes highly significant. Heterogeneity and loss of significance for the overall finding are entirely because of the Xu *et al.* (6) study (based on a "sensitivity analysis;" results not shown) consistent with the larger study size and overall null results.

Dose Response. As indicated above, there is a tendency but not significant evidence in the summary data that the risk for heterozygotes (GA) is intermediate between that of the two homozygous genotypes.

Time Sequence. This criteria is inherent in the germ-line nature of the gene.

Consistency. Given the heterogeneity in the eight published studies, this criteria is not met.

Specificity of Association. There is some evidence for an effect in other tobacco-related cancers (*i.e.,* laryngeal although null in pharyngeal cancer; Ref. 34), whereas a study in lymphoma, a nontobacco-related malignancy, was null (35). Also generally consistent with the postulated mechanism are reported associations with atherosclerosis, stroke, liver fibrosis in hepatitis C (36), Alzheimer's disease, and multiple sclerosis (37), as well as a finding of a relation of the MPO variant to DNA adducts levels in mammary tissue (38). However, all of these associations require additional confirmatory studies.

Summing up these categories of evidence, considering the criteria for causality with regard to MPO and lung cancer, this evidence does not allow a clear conclusion.

Perspectives on Nonconfirmation?

Methodologists from different disciplines have offered diverse perspectives on the reasons why verification of gene-disease associations has been so inconsistent.

Virtually all of the observers note that power has been inadequate, particularly for interactive effects (*i.e.,* gene-environment and gene-gene; Ref. 39). Although population stratification is both a real and theoretical problem, recent work suggests it does not pose a substantial source of bias sufficient to justify the inefficiency and other drawbacks of using family controls in population study settings (40).

Epidemiologists and statisticians have described poor control selection, design flaws, publication bias, disease misclassification ("phenotype" definition in genetics parlance), failure to obtain or poor quality exposure data, misclassification bias, excessive type 1 errors (false-positives), post-hoc and subgroup analysis, and multiple comparisons as the likely culprits (41, 42). Geneticists point to failure to consider the mode of inheritance, genetic and disease heterogeneity, failure to account for the linkage disequilibrium structure of the gene, inadequate characterization of the mutations that account for the genetic abnormalities, and the likelihood that genes identified only account for a small proportion of the variability in risk, *i.e.,* power (43–45).

All of the above points are relevant; however, it may be that additional sources of complexity add to the challenge of identifying the specific genes that account for susceptibility to cancer. Three types of complexity will be considered.

Pheotype Complexity

In explaining failed attempts to find genes that account for major psychiatric conditions such as manic depression and schizophrenia, behavioral geneticists cite the difficulty in defining the phenotype (46). Cancer epidemiologists would call this "disease misclassification," but generally argue that cancer studies enjoy the theoretical advantage of a diagnosis established by pathological tissue examination, rather then one at least historically based on solely clinical criteria. In fact, there is every reason to believe that what we clinically label and traditionally study as "cancer" will not have a one-to-one correspondence with a specific gene (47). Genes may be associated with some early aspect of carcinogenesis (*i.e.,* DNA repair and metabolism), a late feature (*i.e.,* metastasis), a particular histological variant (*i.e.,* lung adenocarcinoma), a vanety of molecular features (i.e. somatic mutations, expression pattern, etc.) or an exposure causally linked to the cancer such as tobacco.

An example is provided that illustrates the last point. Twins studies consistently demonstrate strong heritability for smoking (48, 49). Population studies have tended to rely on statistical adjustments for tobacco use to separate effects of smoking, but given the strong dependency of lung cancers on tobacco, this strategy may not be completely successful. Certain genes may contribute to both smoking and lung cancer, *i.e.,* CYP2A6 is implicated in both the metabolism of nicotine and of carcinogenic nitrosamines (50).

Aside from exposures, component or antecedent conditions may be related to particular genes. Chronic obstructive pulmonary disease and, specifically, emphysema may comprise an independent risk factor for lung cancer (51). Emphysema exhibits a strong familial tendency and at least one gene, *α-1-antitrypsin,* is established as causing a more severe form of this condition (52). Because many of the genes and mechanisms that have been proposed for lung cancer such as GSTM1 (53) and epoxide hydrolase (54) have also been suggested as risk

factors for emphysema, observed associations with either condition may depend on the degree to which the conditions overlap in a particular study setting.

These points illustrate how genes may be associated with exposures or conditions related to the disease of interest, rather than to the disease itself. More generally, whereas early studies of genetic susceptibility in lung cancer focused on the class of genes that controlled metabolism of chemical carcinogens, newly identified genes and pathways involving key features of neoplasia such as inflammation, angiogenesis, DNA repair, drug transport, immune response, metastasis, apoptosis, and others are under active investigation. Studies will need to consider and examine a variety of "endophenotypes" potentially associated with cancer, and expect that specific genes may exhibit a variable and not necessarily consistent degree of association with lung cancer itself.

Gene Complexity

In the late 1980s, as the first RFLP assays began to replace probe drug phenotyping, there was the naïve expectation that variation in function of complex P450s would be explained by a set of simple mutations. This expectation was not unreasonable given that family studies established, for example, for NAT2 and CYP2D6, that metabolic phenotypes segregated in families according to Mendelian expectations (55). With additional study their complexity has become apparent. Both NAT2 and CYP2D6 (>70 mutations) exhibit a variety of common and rare mutations, which ablate activity, increase activity (*i.e.,* gene amplifications), or alter substrate specificity (56, 57). Whereas reasonable assessments of gene activity can be constructed from smaller subsets of the known variants, earlier studies that assessed only one or a few mutations can be seen in retrospect to have had little chance to identify associations. This also suggests that the hope that a given haplotype or SNP can accurately represent gene action, at least for the more complex genes, may not always be fulfilled. Linkage disequilibrium-based search strategies that depend on random SNPs may misclassify and, therefore, miss such gene systems (58, 59).

In the case of MPO, the regulatory site that has been consistently studied may not be the only source of hereditary variation. It is likely that other polymorphisms (some rare) as well as environmental factors contribute to biologically meaningful regulation of MPO (60). It is of interest to note that among rare individuals that lack the gene, increased infections are observed only in the presence of secondary factors (*i.e.,* diabetes). This suggests that parallel pathways may "buffer" activity and that presently unknown combinations of genes may mediate parallel effects. This leads directly to the last aspect of complexity to be considered.

Pathway Complexity

There has been much discussion regarding whether the effects of multiple genes are additive or multiplicative. Postulating simple and straightforward mechanisms is necessary to model statistical expectations, but in actuality nature may be more complex. For example, as noted above, selection has insured that important pathways controlling processes that are critical in neoplasia (cell cycle control, immune response, and DNA repair) are redundant. As an example, P450s have overlapping substrate specificities (61). Many xenobiotics can be metabolized by more than one enzyme albeit with differing kinetic parameters. In addition to the complexity involving specific

substrates, biological endpoints (*i.e.,* cell transformation) can likely be reached in diverse ways. Pathways to phenotypes may be anastomotic, *i.e.,* several pathways lead to the same phenotype with homeostatic and compensating feedback mechanisms between them (62). These considerations imply that biologically meaningful effects may become apparent only when there are sufficient "hits" accrued to a given pathway. The current complex pattern of somatic mutations that characterize specific cancers may come into better focus when pathways are assessed. As a practical application, this would suggest that it will be worthwhile to consider a panel of "Phase I genes" together. Alternatively, a panel of genes that influence a key mediator such as nicotine, nitrosamines or benzo(*a*)pyrene might be considered. Although perhaps exceeding the current analytical capacity and the biochemical and genetic database, integrating projected exposures with kinetic parameters for each relevant genetic variant may provide a more biologically realistic projection of exposure.

Given these issues, we briefly consider the prospects that technology will resolve the question of gene-cancer associations.

What is the Role of Technology in Resolving These Questions?

Accelerating technological progress is characteristic of our era, and there is justifiable enthusiasm that new approaches will improve molecular epidemiology studies and refine the ability to conduct studies to detect associations. High throughput genotyping using ever-smaller quantities of specimens obtained less invasively, increasing availability of SNPs from the Human Genome effort, and improved bioinformatics tools to facilitate information processing are examples that are currently impacting these studies.

There are two different orientations to approaching the problem of complex disease. The traditional approach begins with specific candidate genes such as *MPO* for which some population and mechanistic evidence already exists. Increasingly, technology affords the opportunity to rapidly and efficiently test specific variants against many DNA samples.

The alternative approach, based on the availability of the database and informatics tools to use genetic markers deriving from Human Genome research, and potential for pooling DNA (63) and high throughput genotyping, is to conduct linkage disequilibrium mapping to identify regions that harbor candidate susceptibility genes. This approach may be conducted using an ambitious entire genome search (64) or may be limited to particular regions thought to harbor candidates based on other evidence. There is controversy about the degree to which the extent of linkage disequilibrium (which varies in different population settings and also across the genome) affects this strategy (65, 66). Limiting markers to include a smaller number of representative haplotypes may enhance the ability to identify regions that may harbor new susceptibility genes (67). Given the vast number of available SNP:, and the attendent problem of false positives, it will be important to develop a strategy for SNP selection that takes into account: unferred function, biological plausibility, conservation across species, haplotype setting, and prior population data.

A variety of new technologies will be briefly mentioned that compliment both of these approaches by providing new candidate genes for evaluation and introducing new evidence to support biological plausibility. However, it should be noted that whatever technical approaches are used to suggest new candidates or enhance the evidence for old ones, verification in an

appropriately designed population-based study will remain a benchmark requirement for scientific acceptance and any serious consideration of public health impact.

Some Specific Technologies

Animal models and various model systems provide insight into the genetics of cancer (68). Although there are important intrinsic limitations (different life span, organ and site specificity, complex human exposures, and so forth), the available mouse models should increasingly be a rich source of candidates for human study (69). Expression studies in tumors can provide indications of genes that are under or overexpressed in tumors, and by identifying patterns, suggest fundamental disease redefinition (70). Gene silencing by methylation can alter expression, and so identifying genes that have undergone this epigenetic alteration in critical tissues can identify elements in the transformation pathway. Cytogenetic changes were historically noted as early evidence that the genetic material was a key target in cancer. Newer approaches such as interphase genetics (*i.e.,* fluorescence *in situ* hybridization) have delineated key regions with characteristic alterations in specific cancers. There is great enthusiasm that efforts to isolate protein/peptide markers of cancer will identify "early markers" for cancer and also identify critical genes or pathways.

There is no question that each of these approaches and others have the potential to identify new candidates and to provide new types of mechanistic support. Yet we will ultimately be left with the same task we currently face with regard to MPO, that is, examining the population evidence to establish whether the specific variant of interest is associated with disease.

Some Conclusions

Simple quantitative considerations and our current mechanistic understanding of MPO suggest that the hypothesis of an association of the polymorphic variant with lung cancer remains unproven but still plausible. Specifically, the question of whether the existing positive studies represent a real but small effect exists cannot be resolved with the available data. The broad "complexity" considerations noted above would suggest future studies might consider whether: (*a*) based on its mechanism or substrate specificity MPO might have a relation to some other endophenotype, *i.e.,* inflammation or altered immune function; (*b*) other MPO genetic variants may exist and should be analyzed; and (*c*) other genes known to influence the disposition of benzo(*a*)pyrene and oxidative species should be comprehensively assessed and the possibility of nonadditive/nonmultiplicative relationships among genes in the same pathway explored.

What Can Be Recommended?

The earlier cited recommendations that studies will have to be of optimum design, quality, and size are relevant (71). Population-based designs including both case-control (72) and cohort (73) designs will be key to the identification and verification of candidate susceptibility genes (74). The advantage of the diverse approaches derives from the different opportunities afforded by each. For example, a cohort study can address multiple disease endpoints potentially allowing a simultaneous evaluation of the multiple cancers postulated to be related to MPO as well as other conditions (*i.e.,* Alzheimer's and arteriosclerosis-related diseases). A case-control study might include a more in-depth questionnaire to better define exposure

subgroups or biospecimens to address finer mechanistic questions.

Meta- and pooled analyses that allow a compilation and limited quantitative treatment of existing data, both published and unpublished, will be a useful adjunct. Meta- and pooled analyses have well-known limitations, and while such efforts should be supported, it is unlikely that pooled approaches alone will resolve these questions. They will provide insight and help with subgroup questions (75) and establish a baseline for new investigations.

Application of technological methods listed as well as newer approaches will provide a steady influx of new candidate genes with attractive mechanistic credentials, but confirmation of candidate genes on the population level will always be required. Studies with borderline power reporting new findings will be met with a greater degree of skepticism, whereas journals will find web and other formats to record null studies that heretofore have been considered too low priority to publish (76). Given the recognized obligate role of environmental factors in most common cancers and lung cancer in particular, careful exposure assessment will be essential to evaluate how risks because of genes and environment interplay. Technology will prove to be an asset, particularly when it allows issues of cancer causation to be considered in a broad context.

Final Summary

Epidemiological studies that intend to evaluate genes must be of substantial size. Diverse studies of various (*i.e.,* both case-control and cohort) but sound design will be required, as verification in different populations and exposure settings will be important. Because population differences in both exposure and genes will be considerations, international cooperation will be a priority. Networking studies and consortia will be critical to encourage interdisciplinary and international participation. Leaders must craft incentives into the granting structure to reward such cooperation. If any of the complexity considerations noted earlier apply, the only way small but potentially important effects can be revealed and then confirmed is through cooperation between existing and planned large-scale efforts. From the earliest molecular epidemiology studies, the advantage of close cooperation between population scientists and laboratory investigators was recognized. The large studies and cooperative efforts envisioned will require an infrastructure and level of support capable of generating a rich population data and biospecimen resource to richly reward the new generation of interdisciplinary scientists ready to address these challenges.

References

1. Lander, E., and Schork, N. J. Genetic dissection of complex traits. Science (Wash. DC), *265:* 2037–2048, 1994.

2. Risch, N., and Merikangas, K. The future of genetic studies of complex human disease. Science (Wash. DC), *273:* 1516–1517, 1996.

3. Hirschhorn, J. N., Lohmueller, K., and Hirchhorn, K. A comprehensive review of genetic association studies. Genet. Med., *4:* 45–61, 2002.

4. Dunning, A. M., Healey, C. S., Pharoah, P. D., Teare, M. D., Ponder, B. A., and Easton, D. F. A systematic review of genetic polymorphisms is and breast cancer risk. Cancer Epidemiol. Biomark. Prev., *8:* 843–854, 1999.

5. Feyler, A., Voho, A., Bouchardy, C., Kuokkanen, K., Dayer, P., Hivonen, A., and Benhamou, S. Myeloperoxidase–G463A polymorphism and lung cancer risk. Cancer Epidemiol. Biomark. Prev., 2002.

6. Xu, L-L., Miller, D. P., Zhou, W., Lynch, T. J., Wain, J. C., Su, L., and Christiana, D. C. The myeloperoxidase–463G>A polymorphism does not decrease lung cancer susceptibility in Caucasians. Cancer Epidemiol. Biomark. Prev., 2002.

7. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. Environmental and heritable factors in the causation of cancer. N. Engl. J. Med., 78–85, 2000.

8. Risch, N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. Cancer Epidemiol. Biomark. Prev., 10: 733–734, 2001.

9. Kellermann, G., Shaw, C. R., and Luyten-Kellermann, M. Aryl hydrocarbon hydroxylase inducibiility and brochogenic carcinoma. N. Engl. J. Med., 289: 934–937, 1973.

10. Ayesh, R., Idle, J. R., Ritchie, J. C., Crothers, M. J., Hetzel, M. R. Metabolic oxidation phenotypes as markers of susceptibility to lung cancer susceptibility. Nature (Lond.), 312: 169–170, 1984.

11. Seidegard, J., Pero, R. W., Miller, D. G., and Beattie, E. J. A glutathione transferase in human leukocytes as a marker of susceptibility to lung cancer. Carcinogenesis (Lond.), 7: 751–753, 1986.

12. Lower, G. M., Jr., Nilsson, T., Nelson, C. E., Wolf, H., Gamsky, T. E., Bryan, G. T. N-Acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. Environ. Health Perspect., 29: 71–79, 1979.

13. Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., Devanaboyina, V. S., Nangiu, N. A., and Feng, Y. Molecular Genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol. Biomark. Prev., 9: 29–42, 2000.

14. Marcus, P. M., Vineis, P., and Rothman, N. NAT2 slow acetylation and bladder cancer risk: a metanalysis of 22 case-control studies conducted in the general population. Pharmacogenetics, 10: 115–122, 2000.

15. Bartsch, H., Nair, U., Risch, A., Rojas, M., Wikman, H., and Alexandrov, K. Genetic polymorphisms of CYP genes, alone or in combination, as a risk modifier of tobacco associated cancer. Cancer Epidemiol. Biomark. Prev., 9: 3–28, 2000.

16. Houlston, R. S. Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. Cancer Epidemiol. Biomark. Prev., 8: 675–682, 1999.

17. McWilliams, J. E., Sanderson, B. J. S., Harris, E. L., Richert-Boe, K. E., and Henner, W. D. Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. Cancer Epidemiol. Biomark. Prev., 4: 589–594, 1995.

18. Benhamou, S., Lee, W. J., Alexandrie, A. K., Boffetta, P., Bouchardy, C., Butkiewicz, D., Brockmoller, J., Clapper, M. L., Daly, A., Dolzan, V., Ford, J., Gaspari, L., Haugen, A., Hirvonen, A., Husgafvel-Pursiainen, K., Ingelman-Sundberg, M., Kalina, I., Kihara, M., Kremers, P., Le Marchand, L., London, S. J., Nazar-Stewart, V., Onon-Kihara, M., Rannug, A., Romkes, M., Ryberg, D., Seidegard, J., Shields, P., Strange, R. C., Stucker, I., To-Figueras, J., Brennan, P., and Taioli, E. Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. Carcinogenesis (Lond.), 8: 1343–1350, 2002.

19. Rostami-Hodjegan, A., Lennard, M. S., Woods, H. F., and Tucker, G. T. Meta-analysis of studies of the CYP2D6 polymorphism in relation to lung cancer and Parkinson's disease. Pharmacogenetics, 8: 227–238, 1998.

20. Hoover, R. N. Cancer–nature, nurture or both. N. Engl. J. Med. 343: 135–136, 2000.

21. Rothman, N., Wacholder, S., Caporaso, N. E., Garcia-Closas, M., Buetow, K., Fraumeni, J. R., Jr. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. Biochim. Biophys. Acta, 1471: C1–C10, 2001.

22. Caporaso, N., and Goldstein, A. Cancer genes: single and susceptibility: exposing the difference. Pharmacogenetics, 5: 59–63, 1995.

23. Weed, D. L., and Gorelic, L. S. The practice of causal inference in cancer epidemiology. Cancer Epidemiol. Biomark. Prev., 5: 303–311, 1996.

24. Bradford-Hill A. Principles of Medical Statistics. 1971. London: Oxford University Press.

25. Mallet, W. G., Mosebrok, D. R., and Trush, M. A. Activation of (+)trans-7, 8-dihydrobenzo(a)pyrene to diolepoxides by human polymorphonuclear leucocytes or myeloperoxidase. Carcinogenesis (Lond.), 12: 251–254, 1991.

26. Hoy, A., Leininger-Muller, B., Kutter, D., Siest, G., and Visvikis, S. Growing significance of myeloperoxidase in non-infectious diseases. Clin. Chem. Lab. Med., 40: 2–8, 2002.

27. Masuda, M., Suzuki, T., et al. Chlorination of guanosine and other nucleosides by hypochlorous acid and myeloperoxidase of activated human neutrophils. J. Biol. Chem. 276: 40486–40496, 2001.

28. London, S. J., Lehan, T. A., and Taylor, J. A. Myeloperoxidase Genetic Polymorphism and lung cancer risk. Cancer Res., 57: 5001–5003, 1997.

29. Le Marchand, L., Seifried, A., Lum, A., and Wilkens, L. R. Association of the myeloperoxidase -463G–>a polymorphism with lung cancer risk. Cancer Epidemiol. Biomark. Prev., 9: 181–184, 2000.

30. Cascorbi, I., Henning, S., Brockmoller, J., et al. Substantially reduced eisk of cancer of the Aerodigestive tract in subjects with variant –463A of the myeloperoxidase gene. Cancer Res., 60: 644–649, 2000.

31. Schabath, M. B., Spitz, M. R., Hong, W. K., Delclos, G. L., Reynolds, W. F., Gunn, G. B., Whitehead, L. W., and Wu, X. A myeloperoxidase polymorphism associated with reduced risk of lung cancer. Lung Cancer, 37: 35–40, 2001.

32. Misra, R. R., Tangrea, J. A., Virtamo, J., Ratnasinghe, D., Andersen, M. R., Barrett, M., Taylor, P. R., and Albanes, D. Variation in the promoter region of the myeloperoxidase gene is not directly related to lung cancer riskamong male smokers in Finland. Cancer Lett., 164: 161–167, 2001.

33. Kantarci, O. H., Lesnick, T. G., Yang, P., Meyer, R. L., Hebrink, D. D., McMurray, C. T., and Weinshenker, B. G. Myeloperoxidase -463 (G−>A) polymorphism associated with lower risk of lung cancer. Mayo Clin. Proc., 77: 17–22, 2002.

34. Cascorbi, I., Henning, S., Brockmoller, J., Gephart, J., Meisel, C., Muller, J. M., Loddenkemper, R., and Roots, I. Substantially reduced risk of cancer of the aerodigestive tract in subjects with variant–463A of the myeloperoxidase gene. Cancer Res., 60: 644–649, 2000.

35. Matsuo, K., Hamajima, N., Suzuki, R., Nakamura, S., Seto, M., Morishima, Y., and Tajima, K. No substantial difference in genotype frequencies of interleukin and myeloperoxidase polymorphismsbetween malignant lymphoma patients and non-cancer controls. Haematologica, 86: 602–608, 2001.

36. Reynolds, W. F., Patel, K., Pianko, S., Blatt, L. M., Nicholas, J. J., and McHutchison, J. G. A genotypic association implicates myeloperoxidase in the progression of hepatic fibrosis in chronic hepatitis C virus infection. Genes Immun., 6: 345–349, 2002.

37. Hoy, A., Leininger-Muller, B., Kutter, D., Siest, G., and Visvikis, S. Growing significance of myeloperoxidase in non-infectious diseases. Clin. Chem. Lab. Med., 40: 2–8, 2002.

38. Brockstedt, U., Krajinovic, M., Richer, C., Mathonnet, G., Sinnett, D., Pfau, W., and Labuda D. Analyses of bulky DNA adduct levels in human breast tissue and genetic polymorphisms of cytochromes P450 (CYPs), myeloperoxidase (MPO), quinone oxidoreductase (NQO1), and glutathione S-transferases (GSTs). Mutat. Res., 516: 41–47, 2002.

39. Garcia-Closas, M., Rothman, N., and Lubin, J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. Cancer Epidemiol. Biomark. Prev., 8: 1043–1050, 1999.

40. Wacholder, S., Rothman, N., and Caporaso, N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol. Biomark. Prev., 11: 513–520, 2002.

41. Thompson, J. R. Invited commentary: re: multiple comparisons and related issues in the interpretation of epidemiological data. Am. J. Epidemiol. 147: 801–806, 1998.

42. Little, J., Bradley, L., Bray, M. S., Clyne, M., Dorman, J., Ellsworth, D. L., Hanson, J., Khoury, M., Lau, J., O'Brien, T. R., Rothman, N., Stroup, D., Taioli, E., Thomas, D., Vainio, H., Wacholder, S., and Weinberg, C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am. J. Epidemiol., 156: 300–310, 2002.

43. Wacholder, S., McLaughlin, J., Silverman, D., and Mandel, J. S. Selection of controls in case-control Studies! Principles. Am. J. Epidemiol., 135: 1019–1049, 1992.

44. Morton, N. E., and Collins, A. Tests and estimates of allelic association in complex inheritance. Proc. Natl. Acad. Sci. USA, 95: 11389–11393, 1998.

45. Cardon, L. R., and Bell, J. I. Association study designs for complex disease. Nature Reviews, 2: 91–99, 2001.

46. Sher L. Psychiatric diagnoses and inconsistent results of association studies in behavioral genetics. Med Hypotheses, 54: 207–209, 2000.

47. Chanock, S., and Wacholder, S. One gene and one outcome? No way. Trends Mol. Med., 8: 266–269, 2002.

48. Koopmans, J. R., Slutske, W. S., Heath, A. C., Neale, M. C., and Boomsma, D. I. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. Behav. Genet., 29: 383–393, 1999.

49. True, W. R., Xian, H., Scherrer, J. F., Madden, P. A., Bucholz, K. K., Heath, A. C., Eisen, S. A., and Lyons, M. J. Common genetic vulnerability for nicotine and alcohol dependence in men. Arch. Gen. Psychiatry, 56: 655–661, 1999.

50. Tyndale, R. F., and Sellers, E. M. Genetic variation in CYP2A6-mediated nicotine metabolism alters smoking behavior. Ther. Drug. Monit., 24: 163–171, 2002.

51. Kishi, K., Gurney, J. W., Schroeder, D. R., Scanlon, P. D., Swensen, S. J., and Jett, J. R. The correlation of emphysema or airway obstruction with the risk of lung cancer: a matched case-controlled study. Eur. Respir. J., 19: 1093–1098, 2002.

52. Yang, P., Wentzlaff, K. A., Katzmann, J. A., et al. α1-antitrypsin deficiency allele carriers among lung cancer patients. Cancer Epidemiol. Biomark. Prev., 8: 461–467, 1999.

53. Harrison, D. J., Cantlay, A. M., Rae, F., Lamb, D., and Smith, C. A. Frequency of glutathione S-transferase M1 deletion in smokers with emphysema and lung cancer. Hum. Exp. Toxicol., *16:* 356–360, 1997.

54. Takeyabu, K., Yamaguchi, E., Suzuki, I., Nishimura, M., Hizawa, N., and Kamakami, Y. Gene polymorphism for microsomal epoxide hydrolase and susceptibility to emphysema in a Japanese population. Eur. Respir. J., *15:* 891–894, 2000.

55. Weber, W. W. The genetics of acetylation. *In:* The Acetylator Genes and Drug Response, pp. 32–49. Ann Arbor, MI: 1987. Oxford University Press.

56. Fretland, A. J., Leff, M. A., Doll, M. A., and Hein, D. W. Functional characterization of human N-acetyltransferase 2 (NAT2) single nucleotide polymorphisms. Pharmacogenetics, *11:* 207–215, 2001.

57. Bertilsson, L., Dahl, M. L., Dalen, P., and Al-Shurbaji, A. Molecular genetics of CYP2D6: clinical relevance with focus on psychotropic drugs. Br. J. Clin. Pharmacol., *53:* 111–122, 2002.

58. Hosking, L. K., Boyd, P. R., Xu, C. F., Nissum, M., Cantone, K., Purvis, I. J., Khakhar, R., Barnes, M. R., Liberwirth, U., Hagen-Mann, K., Ehm, M. G., and Riley, J. H. Linkage disequilibrium mapping identifies a 390 kb region associated with CYP2D6 poor drug metabolising activity. Pharmacogenomics J., *2:* 165–175, 2002.

59. Weiss, K., and Clark, A. G. Linkage disequilibrium and the mapping of complex human traits. Trends Genet., *18:* 19–24, 2002.

60. Hoy, A., Tregouet, D., Leininger-Muller, B., Poirier, O., Maurice, M., Sass, C., Siest, G., Tiret, L., and Visvikis, S. Serum myeloperoxidase concentration in a healthy population: biological variations, familialresemblance and new genetic polymorphisms. Eur. J. Hum. Genet., *9:* 780–786, 2001.

61. Shimada, T., Oda, Y., Gillam, E. M., Guengerich, F. P., and Inoue, K. Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes P450 in Salmonella typhimurium. Drug Metab. Dispos., *29:* 1176–1182, 2001.

62. Weiss K. M. Genetic Variation and Human Disease, p. 177. Cambridge University Press. Cambridge, 1993.

63. Risch, N., and Teng, J. The relative power of family-based and case-control designs for linkage disequilibrium studies of comples human diseases I DNA pooling. Genome Res., *8:* 1273–1288, 1998.

64. Uhl, G. R., Liu, Q-R., Hess, W. J., and Naiman, D. Polysubstance abuse–vulnerability genes: genes: genome scans for association using 1004 subjects and 1494 SNPs. Am. J. Hum. Genet., *69:* 1290–1300, 2001.

65. Kruglyak, L., and Nickerson, D. A. Variation is the spice of life. Nat. Genet., *27:* 234–235, 2001.

66. Boehnke M. A look at linkage disequilibrium. Nat. Genet., *25:* 246–247. 2000.

67. Lazarus, R., Klimecki, W., Palmer, L., Kwiatkowski, D., Silverman, E., Brown, A., Martinez, F., and Weiss, S. Single-nucleotide polymorphisms in the interleukin-10 gene: differences in frequencies, linkage disequilibrium patterns, and haplotypes in three United States ethnic groups. Genomics, *80:* 223, 2002.

68. Kimura, S., and Gonzalez, F. J. Applications of genetically manipulated mice in pharmacogenetics and pharmacogenomics. Pharmacology (Basel), *61:* 147–153, 2000.

69. Gregory, S. G., *et al.* A physical map of the mouse genome Nature (Lond.), *418:* 743–750, 2002.

70. Davis, R. E., and Staudt, L. M. Molecular diagnosis of lymphoid malignancies by gene expression profiling. Curr. Opin. Hematol. *9:* 333–338, 2002.

71. Little, J., Bradley, L., Bray, M. S., Clyne, M., Dorman, J., Ellsworth, D. L., Hanson, J., Khoury, M., Lau, J., O'Brien, T. R., Rothman, N., Stroup, D., Taioli, E., Thomas, D., Vainio, H., Wacholder, S., and Weinberg, C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am. J. Epidemiol., *156:* 300–310, 2002.

72. Caporaso, N., Rothman, N., and Wacholder, S. L. Case-control studies of common alleles and environmental factors. J. Natl. Cancer Inst. Monogr., *26:* 25–30, 1999.

73. Langholtz, B., Rothman, N., Wacholder, S., and Thomas, D. Cohort studies for characterizing measured genes. J. Natl. Cancer Inst. Monogr., *26:* 39–42, 1999.

74. Khoury, M. J., and Yang, Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. Epidemiology, *9:* 350–354, 1998.

75. Ioannidis, J. A. P. A., Rosenberg, P. S., Goedert, J. J., and O'Brien, T. R. Commentary: meta-analysis of individual participants data in genetic epidemiology. Am. J. Epidemiol., *156:* 204–210, 2002.

76. Shields, P. G. Publication bias is a scientific problem with adverse ethical outcomes: the case for a section for null results. Cancer Epidemiol. Biomark. Prev., *9:* 771–772, 2000.